DOCUMENT RESUME

ED 155 594 CS 004 088

AUTHOR Ingersoll, Gary M.
TITLE A Non-Recursive Analysis of Antecedents to Reading Achievement.
PUB DATE Mar 78
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Age; *Critical Path Method; Elementary Education; Elementary School Students; Family Income; *Predictor Variables; *Racial Differences; *Reading Achievement; Reading Processes; *Reading Research; *Reading Skills; Sex Differences; Verbal Ability; Vocabulary Development

ABSTRACT

The purpose of this study was to identify the relationship between reading achievement and a set of social and individual variables through the use of a critical path analysis. Data, taken from Cycle II of the National Health Examination Survey, included statistics on a national sample of 7,119 children, ages six through eleven years. Four variables were included in this study: sex, race, age, and reported family income. Analyses reveal that race is not a direct contributor to reading achievement; it is an indirect factor as reflected in family income level and vocabulary development. While verbal ability had a possible reciprocal effect on reading achievement, the impact of reading on verbal ability was negligible. (MAI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED155594

A Non-Recursive Analysis of Antecedents to Reading Achievement

Gary M. Ingersoll
Institute for Child Study
Indiana University
Bloomington, Indiana
47401

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gary M. Ingersoll

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the Annual Conference of
the American Educational Research Association
Toronto, 1978

CS004088

When Willy "the actor" Sutton was asked why he robbed banks, he is reported to have answered, "Because that's where they keep the money." Any social scientist worth his salt would scoff at such a use of the word "because" since the data are at best correlational and besides where was his control group? Clearly Willy Sutton was not an astute social scientist. The use of the word "because" in this case was not appropriate, or it was at least inadequate. Or was it? Recently, a number of significant studies based on clinical or correlational data have freely talked in terms of causality. For example, famous or infamous, the Surgeon General's Report on the causal linkage of cigarette smoking to cancer was based on correlational data; geneticists depend on intergenerational correlations to attribute causal characteristics to gene; the prestigious National Academy of Science has recently concluded that floral carbons are causing a depletion of our earth's ozone layer. How can they do that? Is it possible that other sciences know something about research design that we do not?

In recent years, a number of writers, including Blalock (1964), Simon (1962), Duncan (1966), Kerlinger and Pedhauzer (1975), Land (1969), Heise (1969) and Asher (1976), have proposed that the traditional dependence on "classical" experimental design is dysfunctional in the social sciences

since so much variation is beyond our experimental control and we are left with non-experimental or at best quasi-experimental designs. As an alternative, those authors suggest the use of causal modeling or recursive analysis. More commonly referred to as path analysis, recursive techniques allow assessment of assumed causality through correlational data if those data meet stringent conditions. Susser (1973) has similarly proposed the application of recursive techniques in the health sciences with special reference to epidemiology. In some cases, when a mutual causal relationship is suspected, one might better apply a non-recursive extension to assess that reciprocal relationship.

Reading educators repeatedly call for individualized instruction. However, a stable theoretical scheme within which the effects of individual differences, as antecedents to reading achievement are specified, is lacking. Part of the dilemma in building a causal theory in reading achievement lies in the nature of data that reading researchers are forced to use. We are unable to directly manipulate a subject population to assess causality. Rather, we are left with non-experimental analogs that are traditionally viewed as less than adequate. The purpose of this study is to define the magnitude of causal contributions of a limited set of indi-

vidual and social antecedents to reading achievement first in a recursive model and second, in a non-recursive extension of the first model.

Data Sample

Data for the present study are drawn from Cycle II of the National Health Examination Survey (HES) conducted by the National Center for Health Statistics (NCHS). The HES was divided into three cycles to provide a general medical, physical and psychological profile on American children, youths and adults. Cycle II of the HES includes examinations of 7119 children ages 6-11 taken from a National probabilistic sample. The sample is estimated to be representative of 25 million children (NCHS, 1967) and are currently available for professional use in the form of an integrated micro-data tape.

Variables

Of the wide variety of data available on the Cycle II Data File, a limited subset were selected for the current analysis. Four demographic variables were selected for inclusion in this study. SEX (coded Ø = female, 1 = male), AGE (in 6 month intervals), and RACE (coded Ø = Black; 1 = White) were included as exogenous to the system (32 cases

were lost due to the unclear racial classification as "other"). A measure of socioeconomic status was provided through the use of reported family income (FAMINC) which was presumed, in this study to have a causal contribution due to RACE. Each of the above listed variables has been linked with reading achievement. Racial differences, for example, have long been a focus in reading research. More recently, however, investigators (e.g. Harbert and Bryen, 1976) have speculated that lowered achievement levels among Blacks, rather than reflecting deficient ability, may be a result of incompatible (or different) linguistic schema that interfere with acquisition of reading skills. If this were true, then the partial relationship of RACE to reading achievement should be spurious if controlled through vocabulary ability. Similarly, since vocabulary and reading achievement are lower among lower income Whites, a similar relationship between FAMINC and reading might be expected.

The final two variables included in the scheme were vocabulary (VOC) ability as reflected in the vocabulary subscale of the Wechler Intelligence Scale for Children and reading achievement (READ) as measured in raw score units on the Wide Range Achievement Tests (WRAT).

Anastasiow (1971) has concluded that during "initial reading (the child) will be able to decode only those specific words he has already mastered in his own language and will

recognize only those ideas he has already learned." (p. 20). Selected studies show that improved verbal ability leads to increased achievement, especially in reading (see, e.g., Thompson and Blackwell, 1974).

Analysis I

Recursive Analysis

As described by Wright (1921, p. 557), "The method of path analysis depends on the combination of knowledge of some degree of correlation among the variables in a system with such knowledge as may be possessed of causal relations. In cases in which the causal relations are uncertain, the method can be used to find the logical consequences of any particular hypothesis in regard to them" (my emphasis). Thus, the first task of an investigator is to establish a strong theoretical base within which one may establish generalized relationships in a logical multivariate network. As Wright (1934), p. 193) later warns, however, "path analysis is not intended to accomplish the impossible task of deducing causal relationships from the values of correlation coefficients." Rather, the correlational evidence accures causal characteristics by verification and extension of relationships demonstrated under more controlled, but less generalizable, small experiments and by the logical validity of coefficients within a theoretical network (Mullins, 1972). Path analysis also allows the analysis of causal

attributes of variables not readily amenable to experimental manipulation.

Let us try to develop an example. There is a stable linear relationship between race and reading achievement ($r = .19$). The inference that can be drawn from a correlation coefficient is bidirectional, thus, one could infer that race leads to lowered reading achievement ($A \rightarrow B$) or the converse ($B \rightarrow A$). To develop a causal model we would like to eliminate the latter as a spurious inference. To augment that inference Simon (1962) and Blalock (1964) establish criteria which the relationship must meet before a causal inference may be made:

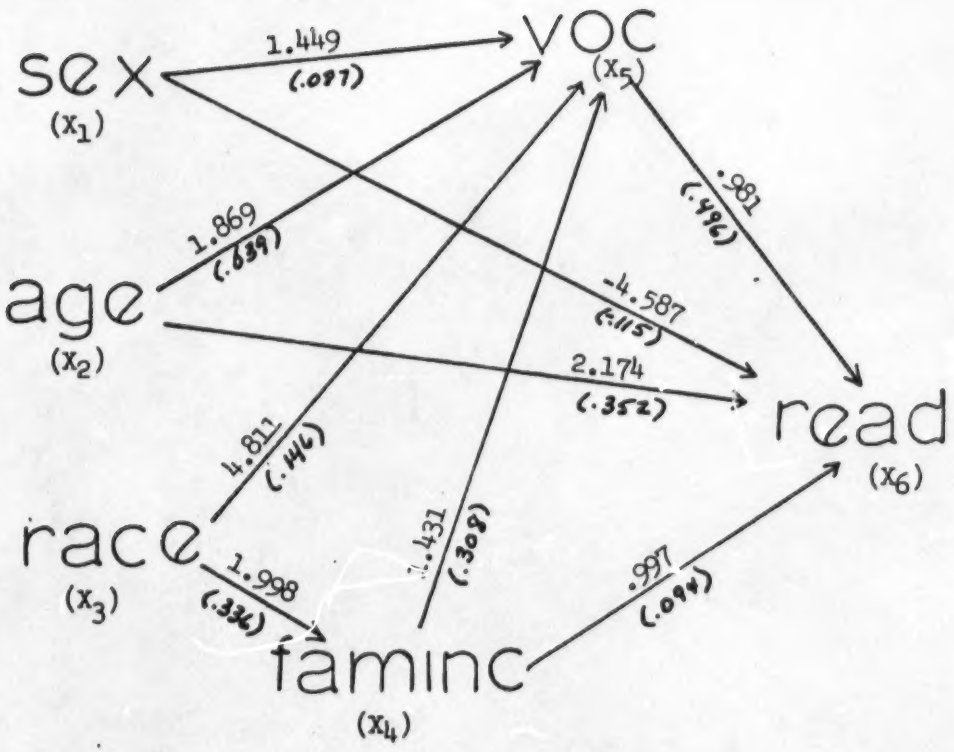
1. There is a linear function that links various variables making up the causal system.
2. One can logically infer a causal impact of one variable on a subsequent variable and that causal relationship is unitary and directional. If A causes B, B cannot cause A. (Non-recursive systems which include such reciprocal relationships are considered later.)
3. One can rule out other confounding factors.

The second assumption is obviously augmented if A can temporally precede B but B cannot temporally precede A. However, because A precedes B does not necessarily imply A causes B. Causality implies not only temporal ordering of A and B but that an alteration of A will lead to a concommitent change

in B. Thus, within finite limits we have established the basis of a causal model between race and reading achievement, albeit a very shallow model, especially when we recognize the potential of other variables related to reading achievement. If we had the luxury of a true experimental setting we could control for such variables by randomization. However, such "true" control is not logically, let alone ethically, feasible. Thus, we control by residualization.

The results of the analysis are presented in the form of a path or arrow diagram with the causal contributions of antecedent variables estimated by path coefficients which are identified through ordinary least squares procedures. When a causal link is reduced to zero then the line or contribution is eliminated from the model. While the path diagram serves as a convenient heuristic, the elements of the model are more accurately reflected in the set of a simultaneous Structural Equations used to define the model.

The results of the recursive analysis are presented as Figure 1. All Path coefficients are unstandardized and least square estimates. The standardized coefficients are presented in parentheses. Note that the direct causal link of RACE on READ was eliminated due to the non-significant contribution (P_{63} = .036). The contribution of FAMINC (P_{64} = .095) while statistically significant ($p < .01$) accounts for only about 1 percent of the variance in reading achievement. VOC and



Structural Equations

$$X_4 = P_{43} X_3 + e_4$$

$$X_5 = P_{51} X_1 + P_{52} X_2 + P_{53} X_3 + P_{54} X_4 + e_5$$

$$X_6 = P_{61} X_1 + P_{62} X_2 + P_{63} X_3 + P_{65} X_5 + e_6$$

Figure 1

Recursive System for Antecedents

to Reading Achievement

AGE remain the most prominant contributors to READ. For each six months of age there is an estimated increase in reading performance of 2.2 raw score units. Likewise each unit of gain in vocabulary growth, as reflected in this measure, shows a concommitant increase of .98 units in WRAT reading scores. On the average we can expect girls to score 4.6 points higher than boys. SEX as a causal contributor to VOC is tenuous. While boys score consistently higher than girls the amount of variance accounted for the relationship is less than 1 percent. The contribution of RACE to reading is at best indirect and when that relationship is controlled for VOC and FAMINC, the relationship disappears and is deemed spurious.

Analysis 2

Non-Recursive Analysis

One might argue that the theoretical scheme depicted in Figure 1 is inadequate since the assumption of a unidirectional effect of vocabulary on reading may not be reasonable. It is possible to speculate that not only does increased vocabulary lead to improved reading but that increased reading might also lead to increased vocabulary. A feedback loop ($X_5 \rightleftarrows X_6$) would substantially alter the interpretation of the model shown in Figure 2.

To estimate the mutual causal parameters of VOC and READ, a Two-Stage Least Squares Solution was used (Hiese, 1975). In a 2SLS solution, each of the non-recursive elements is

regressed on all exogenous variables developing a "prediction" equation for each. The investigator then recomputes the parameters for the structural equations using the "decontaminated" predicted scores as antecedents to each of the non-recursive elements. Path coefficients for the original non-recursive variables are then estimated using ordinary least squares solution. The magnitude of the path coefficients for $\hat{X}_5 \rightarrow X_6$ and $\hat{X}_6 \rightarrow X_5$ would thereby estimate the magnitude of non-recursive effects.

The 2SLS analysis yielded no usable information to confirm a reciprocal loop. The standardized coefficient for $\hat{X}_6 \rightarrow X_5$ was equal to $\emptyset$, while the comparable coefficient for $\hat{X}_5 \rightarrow X_6$ was .307. On the basis of these data then, no substantive evidence exists for a non-recursive model in children.

Discussion

While the results of this analysis are encouraging for the use of recursive analysis in the development of a theory of individual antecedents to reading instruction, they should be viewed with some caution. At a minimum, confirmatory analyses with different data sources and alternative modes of assessing vocabulary growth and reading achievement are in order.

The relationship between vocabulary and reading achievement, although confirming previous assessments of a positive causal relationship, needs to be explored further since the WRAT reading test has a high vocabulary component. One of the obvious limitations to secondary analysis is that an investigator is

restricted to those data included in the original study. While one might argue cogently that better estimates of vocabulary or reading achievement might be used, the argument is not relevant at this level. However, further analysis using this design would benefit from inclusion of other measures, thus increasing generalizability.

One might argue that the state of reading instruction and assessment is at too rudimentary a stage to justify theory building. The counter argument is that one must begin somewhere. The model depicted in the present study establishes a basic theoretical structure which may be added to or modified as new data become available. However, the liklihood of generating a similar data base to Cycle II of the HES in the near future is small due to economic and legal constraints (Hammill, 1977).

The most striking result to be reported from the present study is the failure of RACE to serve as a direct causal contributor to reading achievement. The only causal contributions which may be attributed to RACE are indirect by way of socioeconomic status as reflected in family income (FAMINC) level, and vocabulary development (VOC). Indeed, the direct causal contribution of FAMINC is small. Like RACE, the major impact of FAMINC is mitigated by vocabulary growth. One is tempted to speculate that the primary causal contributor to reading achievement is vocabulary growth and that the traditional depressed

reading performance associated with minorities and lower income groups is more a function of different lexicons than differences in intrinsic ability. If the racial differences in reading achievement were the result of a cognitive deficit, one would expect the causal contribution of RACE to remain, even when controlled for FAMINC and VOC. A linguistic difference model, on the other hand, would suggest that the lowered verbal performance on tasks requiring standard English is primarily a sociocultural characteristic. If one were to eliminate the contamination in reading performance due to sociocultural biases, then achievement differences would not persist. That conclusion, however, is tentative and is contingent upon continued investigation. If the conclusion holds true, it has significant importance for the development of prereading instruction or ancillary instruction intended to facilitate reading achievement among the disadvantaged. As Harbert and Bryen (1976) warn, however, the resolution of the problem is no easy task and while vocabulary growth in standard English serves as a causal contributor to reading achievement, the resolution of the problem may not be simply a case of drilling for improved vocabulary or teaching reading in non-standard English.

Finally, the failure to generate evidence for a non-recursive model which suggests reciprocal causality for reading and vocabulary cannot be taken as the final word. The unique characteristics of the two dependent variables in this study may have led

to the present conclusion of no substantive evidence for a non-recursive system. Further, the non-recursive impact of reading and vocabulary might not be observed until late adolescence or adulthood. However, in light of the present evidence, future investigators might do well to emphasize the recursive nature of the data especially among children.

References

Anastasiow, N. J. Oral language: Expression of thought. Washington, DC: International Reading Association, 1971.

Blalock, H. M. Causal inferences in nonexperimental research. Chapel hill, NC: University of North Carolina Press, 1964.

Coleman, J. S. Effects of school on learning. Review of Educational Research, 1975, 45, 355-386.

Duncan, O. D. Path analysis: Sociological examples. American Journal of Sociology, 1977, 72, 1-16.

Hammill, P. V. V. Height, weight and skinfold measurements. In McKigney, J. I. and Munroe, H. N. (Eds.), Nutrient requirements in adolescence. Cambridge, Mass: M I T Press 1977.

Harber, J. R. and Bryen, D. N. Black English and the task of reading. Review of Educational Research, 1976, 46, 387-405.

Heise, D. R. Causal analysis. New York: John Wiley, 1975.

Mullins, N. C. Theory construction from available materials: A system for organizing and presenting propositions. American Journal of Sociology, 1972, 80, 1-15.

National Center for Health Statistics, Plan, operation, and response results of a program of children's examinations. Vital and Health Statistics, Series 1-No. 5, Public Health Service. Washington, DC: U. S. Government Printing Office, 1967.

Thompson, R. A. and Blackwell, J. M. Verbalized responses to environmental stimuli as stepping stones to decoding written language. Elementary English, 1974, 51, 855-857.

Wright, S. Correlation and causation. Journal of Agricultural Research, 1921, 20, 557-585.

Wright, S. The method of path coefficients. Annals of Mathematical Statistics, 1934, 5, 161-215.